Maintaining legacy data: saving Belfast Harbour (UK) tide-gauge data (1902-2010)

GeoResJ 6 (2015) 65–73

Contents lists available at ScienceDirect

GeoResJ

journal homepage: www.elsevier.com/locate/GRJ

Maintaining legacy data: Saving Belfast Harbour (UK) tide-gauge data (1901–2010)

Joanne Murdy [a], Julian Orford [a,*], James Bell [b]

[a] School of Geography, Archaeology and Palaeoecology, Queen's University, Belfast BT7 1NN, UK
[b] Ashlar Grove, Queensbury, Bradford BD13 2SP, UK

ARTICLE INFO

Article history:
Received 29 September 2014
Revised 31 January 2015
Accepted 9 February 2015
Available online 17 March 2015

Keywords:
Data rescue
Tide gauge data
Analogue–digital conversion
Legacy data
Belfast Harbour (UK)

ABSTRACT

Tide gauge data are identified as legacy data given the radical transition between observation method and required output format associated with tide gauges over the 20th-century. Observed water level variation through tide-gauge records is regarded as the only significant basis for determining recent historical variation (decade to century) in mean sea-level and storm surge. There are limited tide gauge records that cover the 20th century, such that the Belfast (UK) Harbour tide gauge would be a strategic long-term (110 years) record, if the full paper-based records (marigrams) were digitally restructured to allow for consistent data analysis. This paper presents the methodology of extracting a consistent time series of observed water levels from the 5 different Belfast Harbour tide gauges' positions/machine types, starting late 1901. Tide-gauge data was digitally retrieved from the original analogue (daily) records by scanning the marigrams and then extracting the sequential tidal elevations with graph-line seeking software (Ungraph™). This automation of signal extraction allowed the full Belfast series to be retrieved quickly, relative to any manual *x*–*y* digitisation of the signal. Restructuring variably lengthed tidal data sets to a consistent daily, monthly and annual file format was undertaken by project-developed software: Merge&Convert and MergeHYD allow consistent water level sampling both at 60 min (past standard) and 10 min intervals, the latter enhancing surge measurement. Belfast tide-gauge data have been rectified, validated and quality controlled (IOC 2006 standards). The result is a consistent annual-based legacy data series for Belfast Harbour that includes over 2 million tidal-level data observations.

1. Introduction

Church et al. [2] identifies the strategic importance of tide-gauge data sets as a means of determining recent past changes in both mean sea-level and associated water-level variation that relates to storm-generated surges. Many of the sea-level studies identified in Church et al. [2] have utilised sets of sea level and storm data that are less than 30 years in length. As such, forecasts of future change are limited by the quality and length of empirical data that they can access. In effect, uncertainties related to sea-level change rate are significantly reduced when the record quality and length are increased. On this basis, the addition of evidence to this limited set of tide gauges that can be analysed for multi-decade to century scale variation is of considerable strategic importance. Older tide-gauge data sets are dominated by paper records (marigrams), which over time are susceptible to paper degradation and loss. This loss potential means that analogue long-term tide-gauge data records can be legitimately identified as potential legacy data, given that ocean scientists still require these data to be part of continuing tidal data-sets, which are now derived by different mechanisms and stored automatically in digital format. On this basis, the saving and digital abstraction of these paper-based data sets can be considered as of international significance.

Original tide-gauge measurements were based on transferring the actual physical vertical movement of the sea (at one location) into an analogue measurement of proportional vertical distance changing over time. This value was recorded on some form of graph paper, scaled for height (vertical axis) and time (horizontal axis). The traditional method was to measure the tidal rise and fall of a float in a vertical stilling well, connected to an ink pen plotting these movements onto a standard paper graph fixed to a mechanised (originally clockwork) rotational drum. The drum might rotate once every 24 h, or run for 1 or even 2 weeks, after which time a new plot had to be fixed to the drum. These graph paper records were known as marigrams (Fig. 1) and represent the actual source data by which past historical tidal motions were recorded. The statistical analysis of these tidal variations, in particular the

* Corresponding author.
E-mail address: j.orford@qub.ac.uk (J. Orford).

http://dx.doi.org/10.1016/j.grj.2015.02.002
2214-2428/

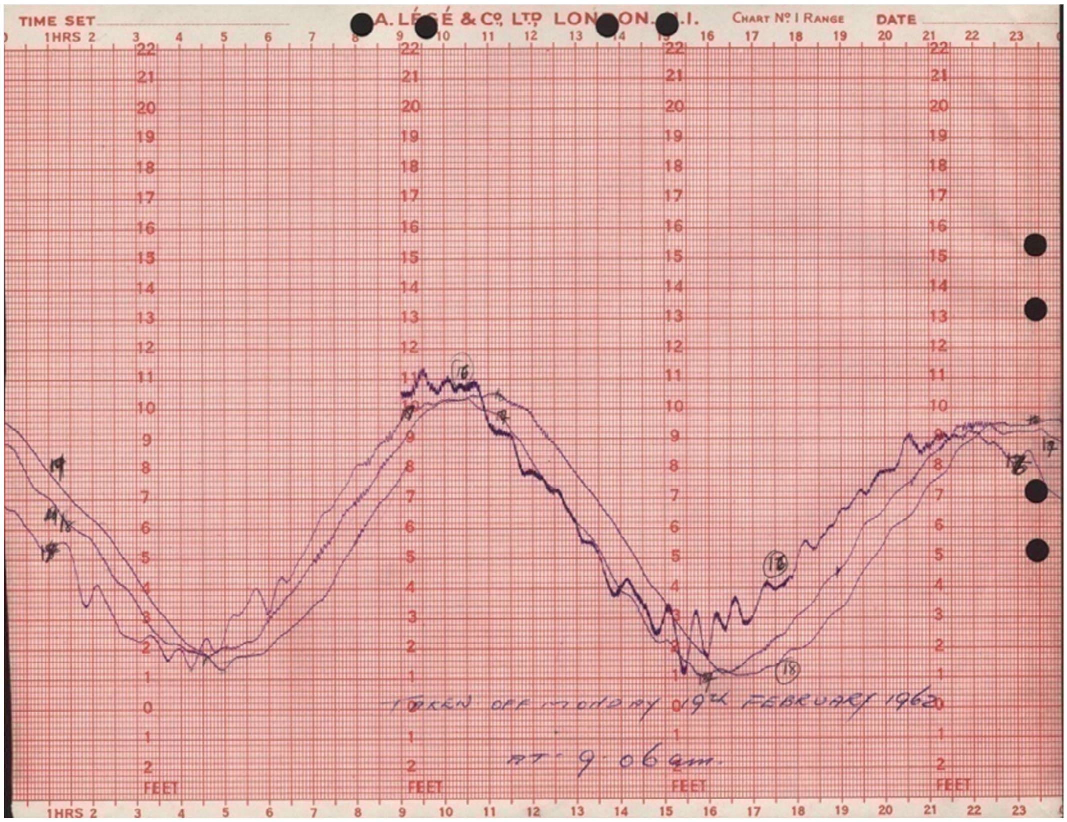

Fig. 1. Original paper marigram from Belfast Harbour showing 3 days tidal recordings (a weekend) with the 'Taken Off' date of 19th February 1962.

mean elevation, is used to determine mean sea-level. Data are normally grouped by chronological year, hence allowing annual mean sea-level determinations to be made.

Although sea-level studies have been traditionally based on a one-hour sampling interval, the development of studies relating to sub-hour processes (e.g. extreme surge analysis) requires the re-sampling of tide-gauge data at higher frequencies. Wang et al. [12] in the context of the latest IPCC forecasts has highlighted the issue of deficient effective studies to support substantive analytical statements of past surge change. This deficiency is principally due to the financial/logistical problems of re-sampling the marigram record at resolutions supportive of surge analysis.

Time has not favoured the storage of paper records. It is evident that new procedures for extracting and thereby saving marigram information on a 'once-only' basis need to be operationalized, since the marigrams are not necessarily going to be available in the future. New data extraction methods in terms of (i) saving the marigram signal and (ii) extracting data at a range of scales from this signal, need to be central in supporting acquisition of analogue tidal-data as legacy data.

The Belfast Harbour Storm Surge Analysis project [5] investigated the periodicities and variations of extreme coastal water levels caused by storm surge activity, identifiable from the Belfast Harbour tide-gauge (Lat: 54° 37′ 06″N; Long 5° 53′ 54″W) in Northern Ireland. This analysis required the improvement of both the quality and length of the Belfast tidal time-series. As a result of this project, the longest tidal time-series currently available in Ireland (1901–2010 and on-going) has been produced. This record adds to the less than 30 tide-gauge records in the world identified by Pugh [10] that exceed 110 years in length (Fig. 2). Entry to this select group reinforces the strategic value of the Belfast data set as a legacy database of considerable time significance. This paper presents the methodology that has established the digital record from the Belfast analogue record. This record is characterised as variable, because of changes in recording procedure used on a daily/weekly basis from four different analogue-based tide-gauges producing the record of water levels in Belfast Harbour during the 20th century. The latest (fifth) Belfast tide gauge is a digital based system (post-1986) producing data to a set format that can be adjusted to meet international standards [1]. The methodology was designed to use either the paper or digital inputs to create a consistent database output for statistical interrogation at IOC standards [5].

The methodology proposed includes (i) digital scanning of marigrams, (ii) extraction of digital tidal elevation at sampling intervals greater than 20 s and (iii) merging and collating variable length marigrams data into annual files. The novel strand in this new methodology is provided by the use of an automated means of digitally acquiring the line graphs (tidal trace) on each marigram (the tidal elevation against time plot), rather than using a standard manual digitisation of the plot, using an *X*–*Y* digitiser. This procedure has produced substantial savings in time whereby a year's worth of marigrams can be reduced to a digital basis in two weeks, compared to two months if using manual digitisation.

2. Methodology

2.1. Data sources, physical condition and datum history of the observed tide-gauge record

Observed tidal information has been recorded by a succession of tide-gauges that were located in Belfast Harbour (Fig. 3). Belfast Harbour tide-gauges were unfortunately never permanent in spatial terms during the 20th century. As a consequence of harbour expansion and re-development over the years, the gauge changed position on several occasions (Table 1), note that TG4

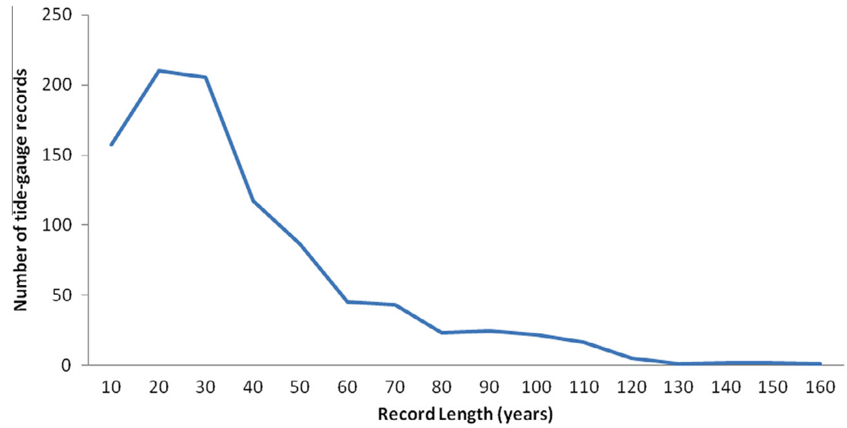

Fig. 2. Effective record length (year) available from the world's tide gauges (number). Note the small number of tide gauges with a record length exceeding 110 years (after Pugh [10]).

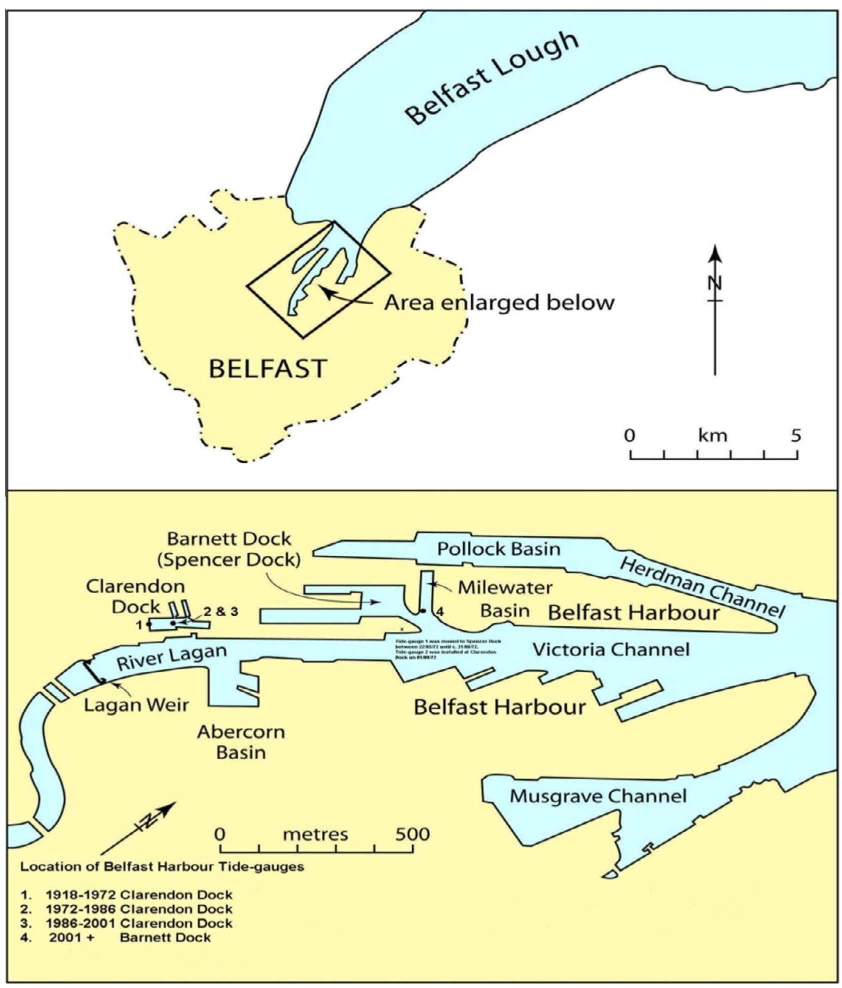

Fig. 3. Belfast Harbour, Northern Ireland. The inset shows the harbour location of the 5 tide gauges used through the 20th century.

and TG5 use the same location. The longest period in which a tide-gauge remained stationary was for 56 years (November 1901 until March 1957), when it was positioned at the SE corner of Clarendon Dock (Table 1). Following harbour expansion and redevelopment, gauge location changed during the 1960's and the 1970's. Finally, the gauge was moved to its present position at Milewater Basin in 1986. Given the importance of tidal analysis being related to an absolute elevation, physical gauge relocation incorporates a levelling process by which the zero level on the gauge is set to a constant elevation. This is known as the tidal reference level. On every relocation occasion, tide-gauge reference level, or Tide Gauge Zero (TGZ) has remained constant, with (TGZ) referring to Belfast Harbour Datum (this level is also British Admiralty Chart Datum (ACD)); and refers to the sill level (entrance limit) of No. 2 Dry dock at Clarendon Dock.

The plethora of observed water level information that exists, outlined in Table 1, is a reflection of changing water-level recording technology and archiving facilities over the last century.

Table 1
Location (shown in Fig. 3) and gauge type in Belfast Harbour 1901–2010.

Date	Tide-gauge	NI grid ref	Description of location	Analogue/digital	Sampling interval	Data years
1901–1957	TG1 Cary-Porter	J344750	SE corner of Clarendon Dock	A	Continuous	47
1957–1972	TG2 Lege (stilling gauge)	J348759	E side of Spencer Dock (Barnett) 31-08-1965	A	Continuous	15
1972–1986	TG3 Lege (stilling gauge)	J344751	SW corner of Clarendon Dock 01-09-1972	A	Continuous	14
1987–2000	TG4 Valeport BTH 700 (marine pressure system)	J350758	Milewater Basin	A/D	60	16
2001 to date	TG5 Aanderra pressure WLTS 3791	J350758	Milewater Basin	D	10	9

From 1986 to 2001, there is an overlap of both digital and paper information, which is an indication of the rapid development of information technology during this period. Prior to this current investigation, only 22% of the Belfast Harbour tidal record existed in a digital format, the remainder of information existed in paper (analogue) form and was held in storage by the Proudman Oceanographic Laboratory at Bidston Observatory, which was the lead institution for analysing tidal variation in the early-mid 20th century in the UK. This analogue record was originally recorded and owned by the Belfast Harbour Commissioners, mainly for port operations and for the production of predicted tides (tide-tables). Annual data collected by Belfast was passed to Bidston for their use and for storage. When the Bidston facilities were relocated to the Department of Oceanography, Liverpool University (2008), the entire record was passed to the British Oceanographic Data Collection (BODC) also located there (now part of the UK National Oceanography Centre). Unfortunately, early years of the Belfast record had been placed in Bidston's cellars and as a consequence were exposed to mildew and other sources of paper deterioration. Fortunately, the UK's Natural Environmental Research Council supported a project with BODC by which much of the early record was chemically treated for mildew, and in effect made it safe for subsequent physical handling. During this re-collation phase, it became apparent that some early 20th century records were missing, while the physical state of some marigrams was approaching near-disintegration, with the additional problem of fading both ink plots and written operator's handling comments on the marigrams. This diminished physical state of the marigrams reinforces the need for digital saving of the tide-gauge data and further underlines the need for development of methods that enhances the legacy attributes of these tidal data sets.

Between 1987 and 2001 TG4 produced a digital measurement of water elevation with paper record of numerical elevation on a 15-min sampling basis. This paper record was eventually discontinued and replaced by an entirely digital tide-gauge recording system in 2002. BODC ceased to receive any further tidal information from the Belfast Harbour Commissioners after 2001, therefore the remainder of the record, which is in digital format was directly acquired from the Belfast Commissioners. It should be noted that the Belfast Harbour tide-gauge is not maintained for sea-level analysis *per se*. In order to produce sea-level data, the raw digital tidal information requires considerable editing and validation before any further analysis can be performed.

Raw digital tide-gauge data were processed following the guidelines outlined by the Intergovernmental Oceanographic Commission [1]. Pugh [9] details the protocols for how observed tide gauge data should be statistically analysed, but these issues are not the concern of this paper.

2.2. Developing the observed tidal record

A fundamental requirement for conventional tide gauge analysis is a good quality, digital observed water-level time series that is based on, at minimum, a 60-min sampling interval. This then allows for the rapid computer-based statistical manipulation and analysis of the data. Given that surge analysis tends to be at a sub-hour scale, the need for a reduced sampling interval is now a major requirement of tidal water-level data series. Lee et al. [3] have illustrated that maximum surge heights calculated from 60 min sampling data considerably underestimated actual surge elevations, compared to analysis based on higher frequency data (10 min). They found that low frequency sampling (equal or greater than one hour) gave a misleading result, particularly for sharply peaked surges operating at periodicities of minutes. This methodology has been designed for acquisition of water level at any sampling period down to the resolution level of the imaging system (20 s), however for this investigation a sampling interval of 10 min was selected, since this was the minimum consistent timing interval appropriate for surge analysis that we could achieve across the majority of the observed Belfast record. The lack of 10 min sampling between 1987 and 2000 meant that the one-hour sampling had to be used as an infill.

The Permanent Service for Mean Sea Level (PSMSL) as part of the NOC (UK) is the principal host to a world collection of digital tidal and sea-level information. PSMSL holds digital data relating to the Belfast Harbour tide-gauge (1918–1963: [11]) but it is based on a Mean Tidal Level (MTL) series; which describes a tidal time-series that comprises the highest and lowest daily tides. http://www.psmsl.org/data/obtaining/stations/219.php). This information had been extended beyond 1963 and subjected to detailed relative sea-level analysis [4,7]. However, this existing MTL series was in an unsuitable format for surge analysis, since the sampling interval of approximately every six hours (half-tidal semi-diurnal periodicity) was too infrequent for effective surge extraction. This data series underlines one of the important principles by which legacy data should be structured: avoid the necessity of repeating data extraction from the marigram *per se* because the accepted standard of sampling has altered. Legacy data needs to be abstracted once from the marigram and then be amenable to supporting past and potential future modes of analysis.

In order to produce a tidal time series that was suitable for surge analysis, it was necessary to extract the observed water level information from all of the original extant paper and digital records of Belfast Harbour tide gauges. The following stages were identified in the design methodology to accomplish this objective.

2.2.1. The collation of existing information

This was the main logistical stage of the process, whereby all of the available analogue records (marigrams) were transferred under loan from the BODC archives to Queen's University, Belfast, for direct scanning. Other sources of data, particularly original raw digital Belfast tide-gauge data was downloaded from PSMSL (http://www.psmsl.org/data/), or provided by the Belfast Harbour Commissioners.

2.2.2. Conversion of analogue record to digital record

The observed tidal record was recorded on the marigram, therefore the marigram becomes the focus by which legacy tidal-data can be defined, examined and stored. It was the duty of the Belfast Harbour Master to ensure that the recording paper for the marigram was regularly changed. As evident from the existing Belfast marigram record, this duty was carried out on a daily and weekly (Monday to Friday) basis, which accounts for longer period records being produced over the weekend (Friday afternoon until Monday morning), as well as over traditional holiday periods (Summer and Christmas). It is only on the weekend marigrams that there is a continuous daily 24-h record of tidal elevation. On the daily marigrams a 24-h record is split over two sheets. Therefore, two separate plots represent a weekday record. Once the Harbour Master removed the marigram, the removal date and time was recorded on each sheet. Additionally, if problems or events were experienced during the recording period, they were often written down on the marigram. This record, in itself, provided a valuable log of information for the production and validation of the observed tidal series. We found on some occasions that these statements were difficult to translate due to fading ink or poor penmanship or simply as the result of physical paper degradation. The ability to retain these operator statements is seen as a valuable supplement to scanning the marigram.

2.2.3. Acquiring data from the marigrams

Over its long history the Belfast tide-gauges have produced approximately 6000 marigrams. Each marigram was scanned in at a minimum resolution of 300 dots per inch (dpi) and saved both as .TIF and .JPG format. This is stage one in saving the tidal record as legacy data. Regardless of subsequent analysis, these files hold an exact copy of the complete marigram plus all associated written commentary. The 300 dpi resolution was chosen to resolve file storage concerns without compromising the quality of the resultant digital archive. Furthermore, an image resolution of 300 dpi allowed for the thinnest pen trace to be easily followed by our graph-line following procedure (the Ungraph™ algorithm). In extreme cases, it was necessary to increase the scanning resolution to 600 dpi, particularly when a thin tidal trace (inked line) had been produced during calm weather periods and/or if a new pen was fitted. Each scanned marigram image was named and can be identified according to its "taken-off" date. The taken-off date simply refers to the date in which the marigram was physically removed from the tide-gauge drum and replaced by new sheet, for example the data from Fig. 1 would simply be named '19-02-1962', saved with the file extension, e.g. .JPG or .TIF.

In 1987 the traditional marigram record was superseded by a new digital system. Due to computer storage limitations at that time, detailed recordings of the observed tides were recorded as printed marigrams using a dot-matrix printer. This early digital tide-gauge system recorded readings at 15-min intervals. These black and white images were scanned in at the lower resolution of 300 dpi.

In 2002 the Belfast Harbour tide-gauge system was updated again following the development of a new vessel tracking system. This 5th and latest tide gauge is entirely digital and does not produce any paper records. Water surface readings are displayed in real time and are logged to file at 10-min intervals. Since this part of the record already exists in digital format, its addition to the existing legacy data system will be discussed later (Section 2.2.6.2).

2.2.4. Digitisation of scanned images using Ungraph™

A plethora of digitising software/methods is commercially available; however the product Ungraph™ v.5 (http://www.biosoft.com/w/ungraph.htm) was found to be the most suitable for defining the marigrams' line-graphs. Ungraph™ displays a user-friendly Windows platform with a built-in help guide. This product is a fast and accurate way to extract the tidal data path as a consistent signal from each marigram. Its principal advantage is that the software automatically allows a cursor to follow the ink plot regardless of line complexity. The line is then re-defined in terms of an X–Y coordinate system.

An example of the sensitivity of Ungraph™ to follow the graph line is shown in Fig. 4A–C. Belfast Harbour was a major shipbuilder until the late 20th century, including the ill-fated Titanic (1912) and its equally ill-fated sister ship, the Britannic (Fig. 4A: launched 26.2.1914, which as a hospital ship hit a mine near Greece on 21.11.1916). Ships of this size (48,158 tons) displaced considerable water volumes during their launch (high spring tide on 28.2.1914), which generated seiching waves in the harbour. These are observed on the marigram (Fig. 4B) for a 15-min period around high tide. The translation of this marigram by Ungraph™ is identified in Fig. 4C where the water level trace, sampled at 60 s (minus the predicted tidal levels), identifies the seiching wave series consequent on Britannic's launch. This reconstruction shows the potential of the 20-s minimum resolution capability of Ungraph™ on the Belfast marigrams. This high-resolution ability shows that the new Belfast data set has the potential for future investigations at temporal scales that are not contemporarily used due to the scarcity of such data.

Although a MATLAB computer program tool kit for digitizing, transforming and validating paper records of sea level variations (NUNIEAU Numerisation des Niveaux d'Eau, http://gcmd.nasa.gov/records/NUNIEAU.html) is available, it was impractical for use on the Belfast Harbour marigrams, given the weekend and holiday marigrams that contained multiple daily records that crossed over each other. Fig. 1 is an example of such a weekend marigram containing 72 h of observed tidal information. This marigram was put on Friday 16th February 1962 at 0900 (GMT) and taken off on Monday at 19th February 1962 at 0906 (GMT), such that the line-graph record crosses over itself on several occasions. Ungraph™ is capable of finding its way across such complex crossovers, by providing a paint option, by which the operator physically marks the area that was to be digitised (see Fig. 5). This is a critical facility when weekend marigrams show overlapping graph lines that had to be individually defined. Ungraph™ also offered a panning option that allowed the zoom in and out of the image to observe the actual tidal trace in greater detail. These utilities allowed for faster and more accurate digitisation of each marigram compared to the NUNIEAU system.

Before each image could be digitised, a co-ordinate system was established by specifying three reference points. These points included the first-point (usually $x = 0$ and $y = 0$), a 'skew' reference point (usually $x = 23$ and $y = 0$) used to correct for physical tilts in the marigram (as set on the recording drum), and finally a second reference control point (usually $x = 23$ and $y = 12$). During digitisation several checks were made of the co-ordinate system to ensure that it was accurate. For example a note was taken of the on/off position and crossing points (at midnight) of each marigram. This is an attempt to identify any problems associated with image distortion or human error.

The last stage of using Ungraph is shown in Fig. 6 where the final scanned images for '13-07-1960.jpg' were digitised to produce 11-07-1960 (0800-2359) (blue line), 12-07-1960 (0000-2359) (red line), and 13-0-1960 (0000-0810) (purple line).

2.2.5. Export of daily records, data management and data back-up

Ungraph digitises the tidal trace as a (streamed) continuous line composed of x, y points. A 24-h tidal trace can contain approximately 6000 points (a tidal value approximately every 20 s). Considering the substantial volume of raw data actually generated by this project, it was necessary to follow a consistent data management and back-up scheme. Each scanned marigram was date-named (e.g. 19-02-1962.jpg) and stored in a folder named according to the month and year it was tidally generated, e.g. the folder Feb_1962 Images, contains all the marigrams that were

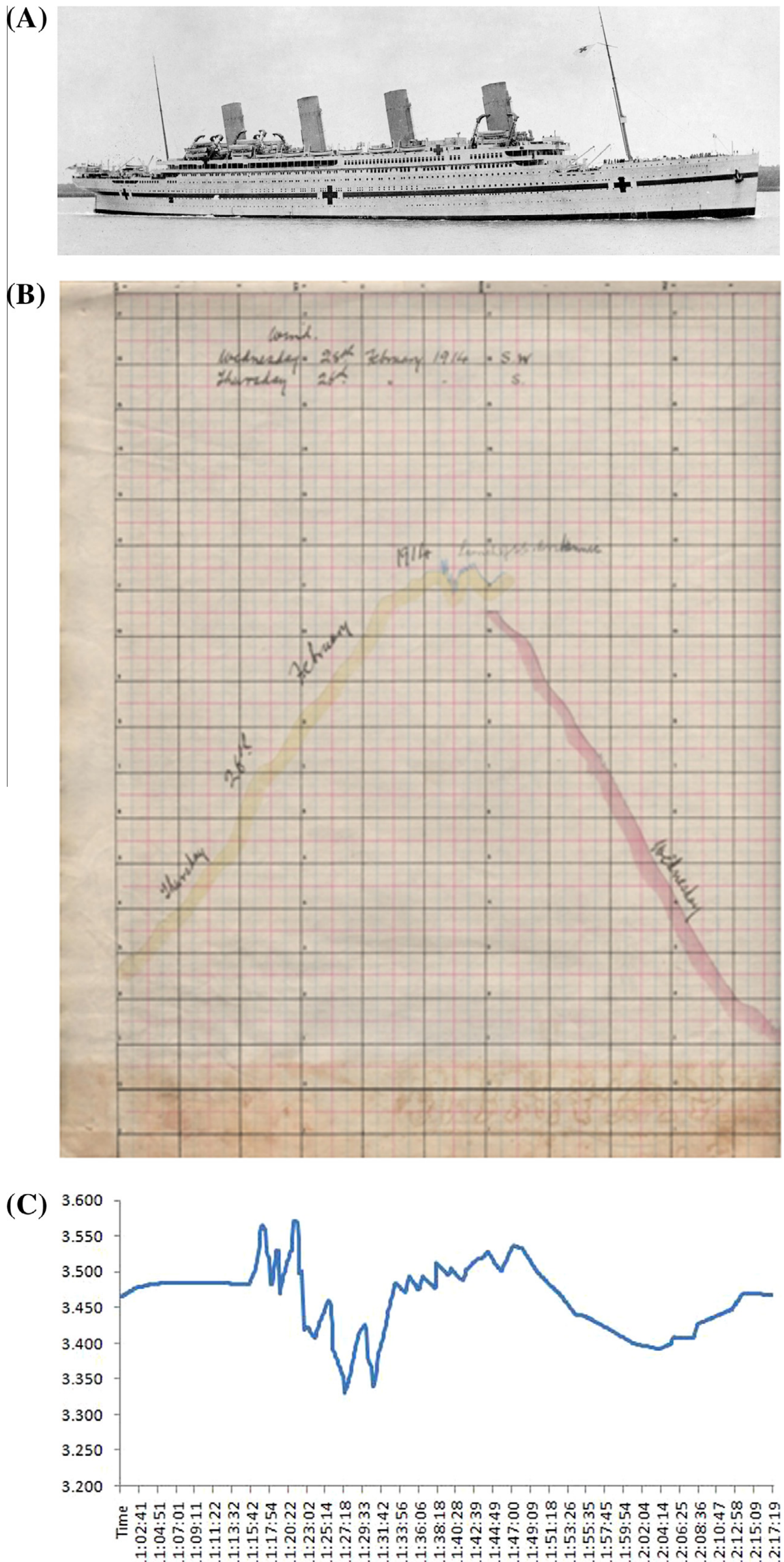

Fig. 4. (A) Britannic as a hospital ship; (B) Launch marigram for 28.2.1914; (C) the reconstructed water level showing seiching waves (vertical scale in meters above chart datum; horizontal scale in minutes).

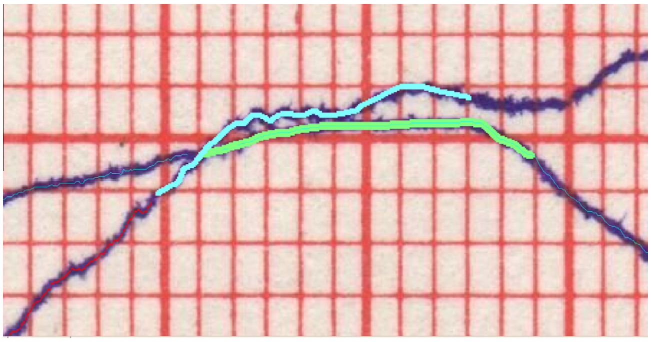

Fig. 5. Enlargement of crossover tidal traces, whereby Ungraph™ allows operator painting of sequential tidal traces.

produced during February 1962. Each month was then put in an annual file named <1962_images>.

Similarly, unedited Ungraph™ files followed a data management procedure as described above. However, in this instance the data was referred to the day the digitised trace was produced. Fig. 1 produced the digitised files referred to as: <16-02-1962(0900-2350)>; <17-02-1962>; <18-02-1962>; and <19-02-1962(0906-2350)>. Where a daily trace was broken up over two marigrams, they were identified according to the time they were produced.

2.2.6. Observe water-level time-series generation

In order to allow for the rapid production of an observed time-series, procedures in Visual Basic for Applications (VBA) were developed, to accurately process and merge the raw Ungraph™ data files of varying time lengths into annual files. These procedures were named Merge&Convert and MergeHYD (program code can be supplied on request). Merge&Convert was developed to deal with the raw files produced by Ungraph™, whereas MergeHYD was developed to process raw tide-gauge report files quickly, specifically those files produced by the most recent fully digital tide-gauge at Belfast Harbour (1.1.2002 onwards). These procedures were designed to reduce the unedited digitised data of varying time lengths, into equal-time intervals. For this investigation, both 10 min and the traditional hourly [9] interval, data were determined in separate data sets (TS1 and TS2, respectively). In effect, over a million water-level observations were determined in this project.

2.2.6.1. Digital data management I: Merge&Convert. Merge&Convert was developed to assist in the editing and validation of unedited digitised data, particularly the information that was produced by Ungraph™. Ungraph™ produced a tidal trace as a (streamed) continuous line composed of x, y points. In order to conduct storm-surge analysis this raw observed data required reduction processing. First, the x-axis, the decimal units need to be converted to time, relating to the 24-h clock. On the y-axis, particularly referring to the pre-decimal years in the UK (before 1972), tidal elevation was converted from feet to metres. Next the code extracts the x and y readings at 10 min intervals. If required, Merge&Convert has the built-in ability to create calendar monthly files or merge monthly files into annual files.

This utility can also identify and deal with any distortion that may have been generated during digitisation, so that individual files can be merged accurately. During processing, the code checks for the correct number of readings, according to the date. Once these series of checks have been performed, the program can locate an already generated predicted tidal data series, copy it onto the worksheet and then calculate the meteorological surge as the difference between observed and predicted tidal level. At this stage

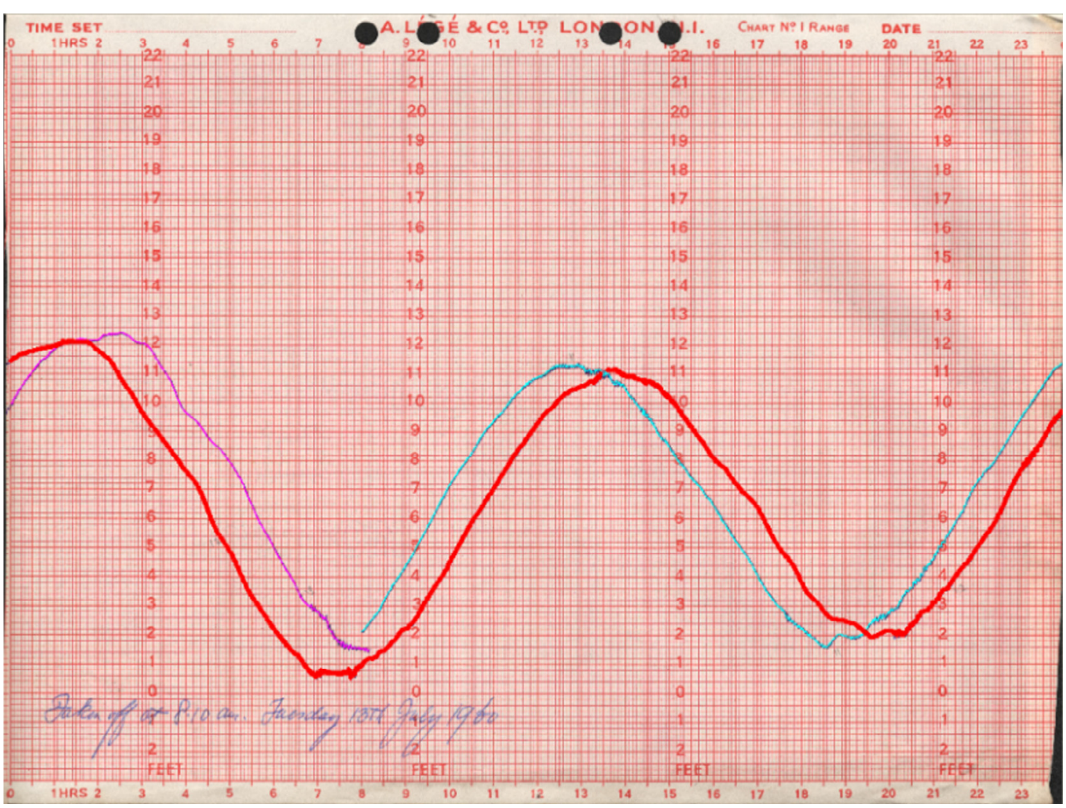

Fig. 6. Example of finished digitised marigram traces. See text for explanation of different coloured traces. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of the project, the generated predicted tidal information is used as a guide to check the resultant digitised time-series.

2.2.6.2. Digital data management II: MergeHYD. MergeHYD was developed to deal with tidal information recorded by the latest Belfast tide-gauge (TG5: Table 1). This software was specifically created to process and add the post-2001 data in a consistent format to the Belfast record. This automatic tide gauge produces a constant stream of information on the tidal elevation to the VTS operations room. Tidal information recorded by TG5 is displayed in real time and are logged to file at a ten-minute interval. Every day this data is compressed and stored as .REP files; which are in a basic ASCII format. The Belfast Harbour Technical Services Department maintain and verify the accuracy of this device on a monthly basis, ensuring that the tide-gauge is calibrated and the monthly tidal readings are collected by the Port Operations Department and forwarded to the UK Hydrographic Office (UKHO).

In order to manipulate the tidal information produced by TG5 for further analysis, the raw tide-gauge data files are imported into MergeHYD where the date, time and water elevation information are merged into monthly or an annual time-series using water levels set to Admiralty Chart Datum (meters) and time set to London (GMT) throughout.

2.3. Tidal analysis and predicted tidal level time series (tidal component)

To extract the meteorological non-tidal residual (storm surge level) from the observed water level time series, a predicted tidal series is required. A detailed, tidal analysis was carried out using MATLAB program "T-Tide" [8], which is capable of performing classical harmonic analysis as well as providing assistance with storm surge analysis (http://www.eos.ubc.ca/~rich/). Up to a total of 67 standard constituents were used in the generation of the predicted tidal series, with shallow water constituents manually added. The extraction of the harmonic constituent set was based on analysis of the 1951 observed water level data. T-Tide was also used to carry out the 18.6 year nodal tide corrections, as well as coloured bootstrap analysis to determine the significance of confidence intervals on the harmonic estimates.

3. Project value

3.1. Record expansion

The original paper-based record, produced by the Belfast Harbour tide-gauge after digitisation has now become the longest empirical tidal time-series available in Ireland. The earliest tidal trace started at 1400 h (Belfast Mean Time) on the 27th of November 1901, whereas the final effective paper record was produced in 2001. This long (88 year) analogue record of the daily tidal activity combined with the continuing contemporary digital record covers a period of 109 complete years (at 31.12.2010). Prior to this project only limited tide-gauge data were available for Belfast. There was some data based on an hourly intervals (TS2_22%), other semi-diurnal high and low tidal levels were available but this was inappropriate for surge analysis. There were no data at 10 min sampling (TS1: 0%) However, following the digital conversion of all the available analogue charts and the merging of the existing digital data, the overall data availability and quality of the resultant time-series can be assessed. Table 2 indicates the percentage of Belfast data availability and quality for both before and after this rescue project. In this instance, any annual data file that is below the 80% availability is considered as being of poor quality for mean sea-level determination *per se*, therefore would be rejected for mean sea-level analysis. However the complete 1902–2010 record can still be used in storm-surge analysis.

Table 2
Summary of Belfast Harbour observed time series availability for 1901–2010, and quality in terms of IOC standard for annual mean sea-level analysis.

	Time-series 1 (10 min interval) (%)	Time-series 2 (60 min interval) (%)
Before this study	0	22
Legacy data set 1901–2010 (100%)	66	78*
Useful annual data (IOC standards) for MSL determination	59	68
Useful surge data	66*	0

* 10 min supplemented by 60 min for 1987–2000.

The Belfast Harbour (TS2) time-series (1-h interval) is now the most complete record, given that only 22% of the potential data-set (annual values) are absent. In comparison, 34% of the more detailed (TS1) (10 min interval) is absent. The quantity of the Belfast Harbour time series further decreases when a mean sea-level quality criteria is applied (i.e. omitting annual records if less than 80% of the annual total is available), consequently, 59% of TS1 and 68% of TS2 remained. However the substantial advance of this rescue program is in the availability of surge analysis data for the first time moving from 0% prior to the study, to 66% availability after this study. This availability now covers most of the 20th century forming a strategic data set by which examination of atmospheric-oceanic forcing in the NE Atlantic can be considered.

The most common reasons for non-availability of data are: tide-gauge failure; the physical destruction of original marigrams; the degradation of the paper due to age (20%); problems of datum control and timing issues (4%); and missing years of marigrams (9%). The latter item relates to 1903, 1907–1909, 1928–1930 and 1933–1934, the records for which are completely missing from BODC's holdings, with no explanation available.

3.2. Scientific expansion

The value of legacy data extraction can be judged by the further development of extreme surge interpretation using the extended data set. The enlarged Belfast data has now been used to specify annual extreme surge through the 20th century [6]. The extension of the data series has allowed the identification of extreme surge variation at a range of periodicities from sub-century to multi-decade to multi-annual scales through the 20th century. The broad scale (sub-century to multi-decades) variation is proposed to be generally consistent with similar broad scale variation in North Atlantic atmospheric forcing as indexed by NAO/AMO interaction.

4. Conclusion

Tide-gauge data can be identified as legacy data, given the transition between analogue and digital recording devices over the century-plus that few tide gauges have been working. There is a strategic need with century-scale objectives for tidal-data continuity, plus the need for re-referencing of century-scaled tidal-level data that requires analogue to digital translation using appropriate legacy methods. Thus past data need to be available for re-analysis despite the shifting mode of measurement and recording type. The output from the Belfast tide gauge (1901 to date), is a characteristic legacy data series. Use of the new methodologies of marigram scanning, automatic graph-line plotting and data merging has translated a variable set of records into a consistent data-base to facilitate standard statistical analysis of both short-term surge and annual sea-level determinations that effectively cover the 20th century. This methodology has in effect eliminated the marigram from future Belfast tide gauge analysis.

Acknowledgements

The Northern Ireland Environment Agency (NIEA), via their Natural Heritage Research Partnership (NHRP) with Quercus, Queen's University, Belfast, funded this project (QU08-09). The British Oceanographic Data Centre, (in particular Prof Philip Woodworth and Dr Liz McDonald) and the Belfast Harbour Commissioners were essential individuals and organisations whose cooperation ensured our access to the various physical and digital forms of Belfast Harbour tide-gauge data accumulated over the 20th century. We should also acknowledge NERC's assistance in association with Prof Woodworth, in funding the rescue, salvage and chemical treatment of the pre-1917 Belfast Harbour marigrams, from the cellars of the Bidston Observatory. We also acknowledge Dr Bjoern Elsaesser (Queen's University, Belfast) assistance in the implementation of MATLAB and T-Tide™ for the predicted tidal elevations.

References

[1] IOC, 2006. Manual on sea level measurement and interpretation. Manuals and Guides 14: An update to 2006, JCOM Tech. Rept No. 31. Intergovernmental Oceanographic Commission of UNESCO, Paris. http://www.psmsl.org/train_and_info/training/manuals/manual_14_final_21_09_06.pdf and http://ftp.wmo.int/Documents/PublicWeb/amp/mmop/documents/JCOMM-TR/J-TR-31/JCOMM-TD-31.pdf.

[2] Church JA, Clark PU, Cazenave A, Gregory JM, Jevrejeva S, Levermann A, Merrifield MA, Milne GA, Nerem RS, Nunn PD, Payne AJ, Pfeffer WT, Stammer D, Unnikrishnan AS. Sea level change. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM, editors. Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press; 2013. p. 1137–216.

[3] Lee J-C, Park K-S, Kwon J-I, Kim S-I. Storm surge calculations using sea level data. Mar Geodesy 2009;32:108–17.

[4] Orford JD, Murdy J, Freel R. 20th century changes in sea level around the north of Ireland. Environment and Heritage Services (Northern Ireland). Research and Development Series, 06/12, p. 106; 2006.

[5] Orford JD, Murdy J. Belfast Harbour surge analysis: Part I. Report prepared by the Natural Heritage Research Partnership, Quercus, Queen's University Belfast for the Northern Ireland Environment Agency. Northern Ireland Environment Agency Research and Development Series No. 11/XX, p. 52; 2011.

[6] Orford JD, Murdy J. Presence and possible cause of sub-century and multi-decadal periodicities in 20th-century extreme coastal surge: Belfast Harbour, Northern Ireland. Global and Planetary Change; submitted.

[7] Orford JD, Smith D, Harman M, Murdy J. Recent sea-level change around the north of Ireland. Unpub. Rept. for the Environment & Heritage Service (NI), QUERCUS Project QU06-11, p. 103; 2007.

[8] Pawlowicz R, Beardsley B, Lentz S. Classical tidal harmonic analysis including error estimates in MATLAB using T_TIDE. Comput Geosci 2002;28:929–37.

[9] Pugh DT. Tides, surges and mean sea-level. Chichester, UK: John Wiley & Sons; 1987. 240 p..

[10] Pugh DT. Changing sea levels, effects of tides, weather and climate. Cambridge University Press 2004. 265 p..

[11] Todd NW. Recent sea-level changes in north and east Ireland. Unpublished M.Sc. Thesis, The New University of Ulster; 1981.

[12] Wang PP, Losada IJ, Gattuso J-P, Hinkel J, Khattabi A, McInnes KL, Saito Y, Sallenger A. Coastal systems and low-lying areas. In: Field CB, Barros VR, Dokken DJ, Mach KJ, Mastrandrea MD, Bilir TE, Chatterjee M, Ebi KL, Estrada YO, Genova RC, Girma B, Kissel ES, Levy AN, MacCracken S, Mastrandrea PR, White LL, editors. Climate change 2014: impacts, adaptation, and vulnerability. Part A: Global and sectorial aspects. Contribution of working group II to the fifth assessment report of the intergovernmental panel on climate change. United Kingdom and New York, NY, USA: Cambridge University Press, Cambridge; 2014. p. 361–409.